Florian Wickelmaier
Forschungsmethoden und Mathematische Psychologie
Fachbereich Psychologie
Universität Tübingen

Übungsaufgaben
Data input and output

Erzeugen Sie ein lauffähiges und kommentiertes R-Skript, in dem Sie die folgenden Aufgaben bearbeiten.

a) Lesen Sie die Datei `ise_singlesubj.txt` in R ein und erzeugen Sie einen Dataframe mit folgender Struktur:

```
> str(dat)

'data.frame':        658 obs. of  7 variables:
 $ trial   : int  1 1 1 1 1 1 1 2 2 2 ...
 $ stimulus: int  1 4 6 3 7 2 5 1 5 7 ...
 $ response: int  1 4 6 3 7 2 5 1 2 7 ...
 $ rt      : num  3.661 1.144 0.888 1.6 1.296 ...
 $ wavfile : Factor w/ 5 levels "fm05i10","fm05i24",..: 4 4 4 4 4 4 4 1 1 1 ...
 $ xcoord  : num  -9.38 -3.04 4.44 2.6 8.58 ...
 $ ycoord  : num  -2.43 2.63 2.19 9.53 -3.48 ...

> levels(dat$wavfile)

[1] "fm05i10"   "fm05i24"   "korean10"  "korean24"  "silence24"
```

Beachten Sie, dass der Dataframe die Variable `random_pre` *nicht* enthalten soll und dass die Faktorstufen von `wavfile` wie angegeben benannt werden.

b) Lesen Sie die erste Zeile von `ise_singlesubj.txt` ein und extrahieren Sie mit `substr()` die Werte der Merkmale `Versuchsperson`, `Praesentationsdauer` und `Abfrage_zufaellig`. Fügen Sie dem vorher erzeugten Dataframe drei Spalten hinzu, die diese Merkmale enthalten.

Die zusätzlichen Spalten sollen folgende Struktur haben:

```
 $ id      : chr [1:658] "vp06-" "vp06-" "vp06-" "vp06-" ...
 $ duration: num [1:658] 1 1 1 1 1 1 1 1 1 1 ...
 $ random  : logi [1:658] TRUE TRUE TRUE TRUE TRUE TRUE ...
```

c) Speichern Sie den Dataframe als Binärdatei mit `save()`. Löschen Sie den Dataframe anschließend aus dem Arbeitsspeicher und laden Sie ihn dann mit `load()`. Überprüfen Sie, ob sich die Struktur verändert hat.

d) Bestimmen Sie den Median der Reaktionszeit für jede Stufe von `wavfile`.

e) Lesen Sie die Dateien `trackdata_subj.csv` und `trackdata_subj2.csv` in R ein und erzeugen Sie jeweils einen Dataframe. Setzen Sie dabei die Argumente von `read.table()` so, dass identische Dataframes entstehen. Überprüfen Sie dies mit `identical()`. Beachten Sie, dass in den Dateien fehlende Werte unterschiedlich kodiert sind.

f) Öffnen Sie die Datei `indometh.xlsx` in Excel und speichern Sie diese, ohne sie in Excel zu verändern, als tabstopp-getrennte Textdatei `indometh.txt`. Lesen Sie die Textdatei in R ein. Der resultierende Dataframe soll folgenden Output liefern:

```
> summary(dat)

   Subject         time           conc
Min.   :1.0   Min.   :0.250   Min.   :0.0600
1st Qu.:2.0   1st Qu.:0.750   1st Qu.:0.1150
Median :3.5   Median :2.000   Median :0.3600
Mean   :3.5   Mean   :2.919   Mean   :0.6249
3rd Qu.:5.0   3rd Qu.:5.000   3rd Qu.:0.8650
Max.   :6.0   Max.   :8.000   Max.   :2.7200
              NA's   :1       NA's   :11
```